Vowel identification at high fundamental frequencies in minimal pairs

Friedrichs, Daniel ; Maurer, Dieter ; Suter, Heidy ; Dellwo, Volker

Abstract: The question of vowel intelligibility as a function of F0 is still a matter of debate. Above all concerning vowel sounds produced at F0s exceeding vowel related statistical F1 in citation-form words ('oversinging' F1), it is unclear whether vowel category perception inevitably shifts towards the neighboring category with a higher F1 or can be maintained in such cases. In this study, we tested listeners' perception of the long German vowels /i-y-e-ø- -a-o/ produced by a trained female speaker in the context of minimal pair words (/l-V-gen/) at nine F0-levels between 220 and 880 Hz. Results showed that vowel identification was maintained > 80% up to F0=740 Hz for /e-ø- / and up to F0=880 Hz for /i-y-a-o/. Thus, vowel identification could be maintained in cases of F0 significantly exceeding F1. The role of neighboring vowels, vowel duration, and other productional and acoustical aspects relevant for vowel perception at different F0s is discussed.

Originally published at:
Friedrichs, Daniel; Maurer, Dieter; Suter, Heidy; Dellwo, Volker (2015). Vowel identification at high fundamental frequencies in minimal pairs. In: 18th International Congresses of Phonetic Sciences, Glasgow, Scotland, 10 August 2015 - 14 August 2015. s.n., 1-4.

VOWEL IDENTIFICATION AT HIGH FUNDAMENTAL FREQUENCIES IN MINIMAL PAIRS

Daniel Friedrichs [1], Dieter Maurer [2], Heidy Suter [2], Volker Dellwo [1]

[1] Phonetics Laboratory, Department of Comparative Linguistics, University of Zurich
[2] Institute for the Performing Arts and Film, Zurich University of the Arts
daniel.friedrichs@uzh.ch, dieter.maurer@zhdk.ch, heidy.suter@zhdk.ch, volker.dellwo@uzh.ch

ABSTRACT

The question of vowel intelligibility as a function of F_0 is still a matter of debate. Above all concerning vowel sounds produced at F_0s exceeding vowel-related statistical F_1 in citation-form words ('oversinging' F_1), it is unclear whether vowel category perception inevitably shifts towards the neighboring category with a higher F_1 or can be maintained in such cases. In this study, we tested listeners' perception of the long German vowels /i-y-e-ø-ɛ-a-o/ produced by a trained female speaker in the context of minimal pair words (/l-V-gen/) at nine F_0-levels between 220 and 880 Hz. Results showed that vowel identification was maintained > 80% up to F_0=740 Hz for /e-ø-ɛ/ and up to F_0=880 Hz for /i-y-a-o/. Thus, vowel identification could be maintained in cases of F_0 significantly exceeding F_1. The role of neighboring vowels, vowel duration, and other productional and acoustical aspects relevant for vowel perception at different F_0s is discussed.

Keywords: Speech perception, vowels, F_0, formants

1. INTRODUCTION

Several studies indicate that vowel intelligibility is compromised when the fundamental frequency (F_0) significantly exceeds the first formant frequency (F_1) in terms of both speaker-specific and statistical F_1 in citation-form words (the latter produced at $F_0 \approx 220$ Hz). Early support for this view goes back to self-experiments by Helmholtz [8] who observed that the vowel /u/ shifts towards /o/ if the corresponding sound is produced at F_0 exceeding 175 Hz. In a more detailed study, Howie and Delattre [11] investigated the intelligibility of the five English and four French vowels /i, e, a, o, u/ and /y, ø, ɛ̃, ɔ̃/ sung by a baritone and a soprano (legit style) in isolation (hereafter V condition) at different levels of F_0 (ranges of F_0 = 132–396 Hz for the baritone, 264–1056 Hz for the soprano). They found that the identification performance of the listeners generally decreased when F_0 exceeded F_1 of a vowel in question. Hollien et al. [10] studied the perception of the three corner vowels /a, i, u/ produced in V condition by 18 professional male and female singers (legit and musical-theatre styles as well as singing teachers, ranges of F_0 = 62–554 Hz for male singers, 165–1319 Hz for female singers). They found that when F_0 of a sound exceeded F_1 of a back or front vowel, its perception shifted to the back or front vowel with the next higher F_1 and then to /a/, (i.e., /i/ shifted progressively to /I/, /ɛ/ and then /a/, and /u/ shifted to /U/, /ɔ/, /ʌ/, and then /a/, respectively). Deme [4] investigated the perception of these three corner vowels produced by a single professional soprano singer (legit style) in V condition as well as in consonantal context, i.e., CVC condition. She found further support for this view for both production conditions. Identification rates dropped below 50% at $F_0 \geq 260$ Hz for /i/ and ≥ 350 Hz for /u/, while the identification rate of /a/ remained > 80% up to F_0 = 988 Hz for unaltered V and CVC conditions, and > 60% up to the same F_0 level for isolated vowels with the onset of voicing removed.

In his attempt to define an upper limit for F_0 of identifiable vowels in singing (legit style), however, Sundberg [22, 23] takes a more prudent stand. Searching for the highest percentage of correct identifications observed in various investigations of sung vowels [1, 7, 14, 16, 19, 20], he concluded for a possible identification ≥ 80% of all vowels up to $F_0 \approx 500$ Hz although this frequency exceeds substantially F_1 of vowels such as /i, y, u/. As an explanation, he refers to pitch-dependent formant frequencies in singing, above all used by female singers, and states that, in such a singing technique, the decrease of vowel intelligibility is limited while loudness is gained [21, p. 129]. Moreover, referring to Smith and Scott [17], he indicates possible vowel identification for sounds at even higher levels of F_0, above all when produced in CVC condition [23, p.87], and referring to Gottfried and Chew [6], he points out the impact of a raised larynx for the production of intelligible vowel sounds.

Smith and Scott [17] indeed reported results of a perceptual test of the front vowels /i, I, ɛ, æ/, produced by a soprano in legit style as well as with raised larynx, which showed an identification rate of 70% for all vowels up to F_0 = 880 Hz in V condition with raised larynx and of 70–76% for all vowels up to F_0 = 1108 Hz in CVC condition in legit style as well as with raised larynx. In a recent study of vowel perception in the singing and speaking in Cantonese

Opera style, Maurer et al. [12] reported identification rates ≥ 80% up to $F_0 \approx$ 820–860 Hz for the front and back vowels /i, a, ɔ, u/ produced as syllables (C)V or (C)V:S. In line with this, yet concerning the perception of vowels at high F_0s produced by untrained speakers, Maurer and Landis [13] reported high identification rates ≥ 90% for all of the five long German vowels /i, e, a, o, u/ produced by children in V condition up to $F_0 \approx$ 660 Hz, and for the corner vowels /i, a, u/ up to $F_0 \approx$ 840 Hz.
Thus, the results in the literature are inconsistent and we are left with the question whether or not vowel intelligibility is substantially compromised at F_0s significantly exceeding typical F_1 values and, therefore, an increase of F_0 is accompanied by perceptual shifts from vowels with low F_1 towards vowels with medium and high F_1. The present study addresses this question by means of an investigation of the identifiability of the long German vowels /i, y, e, ø, ɛ, a, o/ produced by a female speaker (professional musical-theatre singer) in the context of minimal pair words (/l-V-gen/) at nine levels of F_0 in the range of 220–880 Hz and perceptually tested in a listening test involving 28 subjects. Hereafter, the vowels are separated into three subgroups, the front vowels /i, y, e, ø, ɛ/, the vowel /a/, which was produced by the speaker within the range of /a-ɑ/ (no front-back classification applicable), and the back vowel /o/. The vowel /u/ has not been included because the word *lugen* is not a commonly known and used lexical unit in the German language.

2. METHODS

2.1. Subjects

A group of 28 Swiss German native listeners (all students at the University of Zurich; 15 female, 13 male; mean age = 23.1, sd = 1.5) participated in the experiment. None of them reported any kind of hearing impairments.

2.2. Stimuli and apparatus

A female speaker (age = 33; Swiss German native speaker, professional musical-theatre singer) produced the German vowels /i, y, e, ø, ɛ, a, o/ in /l-V-gen/ context at F_0 of 220, 440, 587, 659, 699, 740, 784, 831 and 880 Hz. Digital recordings (44.1 kHz sampling rate, 24-bit resolution) were made in a noise-controlled room at the University of Zurich using a cardioid condenser microphone (Sennheiser MKH 40 P48 with pop shield) and an audio interface (Fireface UCX) connected to a PC. The speaker-microphone distance was 30 cm.
For each of the F_0s investigated, the speaker was instructed to produce the vowels in word pairs as minimal pairs within two sets of vowel contrasts, front vowels and /a/, and the back vowel /o/ and /a/, in AB and BA order. Thus, all vowels were contrasted with each other within the sets of /i, y, e, ø, a/ and /o, a/ in the two possible orders of the words in a pair, e.g., *liegen* vs *lügen, lügen* vs *liegen, liegen* vs *legen, legen* vs *liegen* etc.
Piano notes were presented as reference sounds to the speaker via loudspeaker immediately preceding the production.
Listening to the utterances (first and second author), for each vowel and each level of F_0, the word token that appeared to manifest the optimal correspondence between the intended and the perceived vowel category was chosen for further investigation. Thus, for each level of F_0, each of the seven vowels was represented by one /l-V-gen/ token (N = 63; 7 words * 9 F_0s).
Mean F_0 was calculated for 250 ms in the middle of a vowel sound in Praat [3] using the algorithm described in [2]. A maximum deviation from the intended F_0 of 1.9% was found.

2.3. Procedure (listening test)

Single stimuli /l-V-gen/ were randomly presented to the participants of the listening test via closed dynamic headphones (Beyerdynamic DT 770 Pro) in a small and noise-controlled room. On a computer screen, buttons labeled with the seven investigated words were randomly arranged in a circle to account for a potential directional bias of the listeners. Above the response buttons, the sentence *Welches Wort hörst Du?* (*Which word do you hear?*) could be read. When listening to a word, subjects were asked to assign one of the words presented on the screen (seven-alternative forced choice word identification task). After a response, the next stimulus was presented with a delay of 1 sec.

2.4. Data analysis

To approximate the speaker-specific F_1 at a level of F_0 comparable to statistical F_1 in citation-form words, mean F_1 values were calculated for the steady-state mid 250 ms of the vowels produced at F_0 = 220 Hz, using Praat (Burg algorithm for LPC, default settings for female speakers). Calculated formant frequency values were double-checked on the basis of the respective spectrograms.
Total duration of all vowel sounds from onset to offset was measured in Praat with the help of wideband spectrograms. Durations were averaged for each vowel category to investigate the influence of durational information on vowel identification at higher F_0s.

Identification rates (hereafter ID rate) in terms of the correspondences between the intended and the perceived vowels were determined for each F_0. Referring to Sundberg [23, p. 87], an ID rate > 80% was considered as accurate vowel intelligibility, and only the cases with a lower rate were investigated in more detail.

3. RESULTS

Table 1 shows a comparison of the statistical mean F_1 in citation-form words ($F_{1[stat]}$) obtained by Pätzold and Simpson [15] and the mean F_1 estimations for each vowel produced by the investigated speaker in /l-V-gen/ context at $F_0 = 220$ Hz ($F_{1[speaker]}$). The values for /i, e, ø, o, a/ are in good accordance. However, $F_{1[speaker]}$ is substantially lower than $F_{1[stat]}$ for /y/. Since Pätzold and Simpson do not report values for the long vowel /ɛ/, and because no reliable estimation of $F_{1[speaker]}$ in terms of a correspondence of LPC-values and spectrogram for /ɛ/ was possible, no corresponding F_1 were considered for this vowel.

Vowel	$F_{1[stat]}$ (Hz)	$F_{1[speaker]}$ (Hz)
i:	329	367
y:	342	240
e:	431	442
ø:	434	421
o:	438	416
ɛ:	--	--
a:	779	865

Table 1: Mean statistical F_1 values ($F_{1[stat]}$) for Standard German vowels and mean F_1 estimations (mid 250 ms) for the vowels produced by the female speaker at $F_0 = 220$ Hz ($F_{1[speaker]}$).

The listeners' identification performance is shown in the confusion matrices in Figure 1 (one matrix for each F_0 level). With the exception of /ø/ at $F_0 = 587$ Hz, all ID rates proved to be ≥ 23/28, i.e., > 80% up to F_0 of 740 Hz. For the vowels /i, y, o, a/ this even holds true up to F_0 of 880 Hz.

***F_0* = 220 Hz**

	i:	y:	e:	ø:	o:	ɛ:	a:	
i:	28	0	0	0	0	0	0	28
y:	0	28	0	0	0	0	0	28
e:	0	0	25	0	0	3	0	28
ø:	0	0	0	28	0	0	0	28
o:	1	1	0	0	26	0	0	28
ɛ:	0	0	0	0	0	27	1	28
a:	0	1	0	0	0	0	27	28
	29	30	25	28	26	30	28	196

***F_0* = 440 Hz**

	i:	y:	e:	ø:	o:	ɛ:	a:	
i:	27	0	1	0	0	0	0	28
y:	0	28	0	0	0	0	0	28
e:	0	0	26	1	0	1	0	28
ø:	0	0	0	28	0	0	0	28
o:	0	0	0	0	27	0	1	28
ɛ:	0	0	0	0	0	28	0	28
a:	0	0	0	0	0	0	28	28
	27	28	27	29	27	29	29	196

***F_0* = 587 Hz**

	i:	y:	e:	ø:	o:	ɛ:	a:	
i:	28	0	0	0	0	0	0	28
y:	0	27	0	0	1	0	0	28
e:	0	0	26	0	0	2	0	28
ø:	0	1	1	10	0	15	1	28
o:	0	0	1	0	27	0	0	28
ɛ:	0	0	0	0	0	28	0	28
a:	0	0	0	0	3	0	25	28
	28	28	28	10	31	45	26	196

***F_0* = 659 Hz**

	i:	y:	e:	ø:	o:	ɛ:	a:	
i:	28	0	0	0	0	0	0	28
y:	0	28	0	0	0	0	0	28
e:	0	0	24	1	0	2	1	28
ø:	0	3	1	24	0	0	0	28
o:	1	0	0	0	27	0	0	28
ɛ:	0	0	0	0	0	27	1	28
a:	0	0	0	0	4	0	24	28
	29	31	25	25	31	29	26	196

***F_0* = 699 Hz**

	i:	y:	e:	ø:	o:	ɛ:	a:	
i:	28	0	0	0	0	0	0	28
y:	0	27	0	0	0	1	0	28
e:	1	0	23	0	0	4	0	28
ø:	0	0	0	28	0	0	0	28
o:	0	0	1	0	27	0	0	28
ɛ:	0	0	0	0	0	28	0	28
a:	0	0	0	0	3	0	25	28
	29	27	24	28	30	33	25	196

***F_0* = 740 Hz**

	i:	y:	e:	ø:	o:	ɛ:	a:	
i:	26	0	0	0	0	2	0	28
y:	0	28	0	0	0	0	0	28
e:	0	2	23	1	0	2	0	28
ø:	0	2	0	22	3	1	0	28
o:	0	0	0	0	28	0	0	28
ɛ:	0	0	1	1	0	23	3	28
a:	0	0	1	0	1	0	26	28
	26	32	25	24	32	28	29	196

***F_0* = 784 Hz**

	i:	y:	e:	ø:	o:	ɛ:	a:	
i:	28	0	0	0	0	0	0	28
y:	1	26	1	0	0	0	0	28
e:	4	0	15	4	0	5	0	28
ø:	0	1	1	26	0	0	0	28
o:	0	0	0	0	28	0	0	28
ɛ:	0	0	0	0	0	16	12	28
a:	0	0	0	0	1	1	26	28
	33	27	17	30	29	22	38	196

***F_0* = 831 Hz**

	i:	y:	e:	ø:	o:	ɛ:	a:	
i:	28	0	0	0	0	0	0	28
y:	0	28	0	0	0	0	0	28
e:	0	0	17	0	0	11	0	28
ø:	0	0	9	6	0	12	1	28
o:	0	1	0	0	27	0	0	28
ɛ:	0	0	0	0	1	16	11	28
a:	0	0	0	0	1	0	27	28
	28	29	26	6	29	39	39	196

***F_0* = 880 Hz**

	i:	y:	e:	ø:	o:	ɛ:	a:	
i:	28	0	0	0	0	0	0	28
y:	0	28	0	0	0	0	0	28
e:	0	0	14	2	1	11	0	28
ø:	0	2	1	18	6	1	0	28
o:	0	0	0	0	28	0	0	28
ɛ:	0	0	0	0	0	24	4	28
a:	0	0	0	0	0	1	27	28
	28	30	15	20	35	37	31	196

Figure 1: Confusion matrices showing intended vowels (column 1) versus perceived vowels (column 2–8) for all F_0s investigated. Number of listeners = 28. In a single matrix, the bottom row shows the total number of vowel category responses.

In addition to the case of /ø/ at $F_0 = 587$ Hz, substantial confusions (ID rates < 80%) and corresponding perceptual shifts towards other vowel categories occur for the utterances of the vowels /e, ø, ɛ/ in the F_0-range of 740–880 Hz (see Table 2).

V_{int}	F_0	i:	y:	e:	ø:	o:	ɛ:	a:	oriented confusion
ø:	587	0	1	1	**10**	0	15	1	ø:-ɛ:
ø:	740	0	2	0	**22**	3	1	0	
e:	784	4	0	**15**	4	0	5	0	
ɛ:	784	0	0	0	0	0	**16**	12	ɛ:-a:
e:	831	0	0	**17**	0	0	11	0	e:-ɛ:
ø:	831	0	0	9	**6**	0	12	1	ø:-ɛ:,ø:-e:
ɛ:	831	0	0	0	0	1	**16**	11	ɛ:-a:
e:	880	0	0	**14**	2	1	11	0	e:-ɛ:
ø:	880	0	2	1	**18**	6	1	0	

Table 2: Intended vowels at the levels of F_0 for which ID rates dropped below 80% (< 23/28 correct responses). Strong oriented confusions (perceptual vowel category shifts > 50% of the number of correct responses) are displayed on the right.

For /e/, ID rate was < 80% for F_0 of 784, 831 and 880 Hz. However, a strong and oriented perceptual shift was only found for F_0 of 831 and 880 Hz in terms of a shift towards /ɛ/. This vowel can be considered as related to the vowel with the next higher F_1 (see, e.g, [9]).

For /ø/, ID rate was < 80% for F_0 of 587, 740, 831 and 880 Hz. Strong and oriented shifts were only found for F_0 of 587 and 831 Hz towards /ɛ/ and /e/, respectively. /ɛ/ can again be considered as related to the vowel with the next higher F_1. However, this is not the case for /e/, for which the vowel-related differences in the formant patterns at F_0 in citation-form words concern F_2 and F_3 (see [15]).

For /ɛ/, ID rate was < 80% for F_0 of 784 and 831 Hz. Strong and oriented shifts were found for both levels of F_0 towards /a/, i.e., to the vowel with the highest F_1.

Mean duration of the vowel sounds was 622 ms (sd = 99 ms; range = 430–868 ms). One-way ANOVA revealed significant difference ($F(6,56) = 2.34$, $p < .05$) in sound duration of the seven vowels investigated. Tukey's HSD tests only revealed a significant difference in sound duration for /y/ and /a/ ($p = .04$), and /y/ and /o/ ($p = .03$). No significant difference could be found in sound duration for all vowel pairs of /i, o, e, ø, ɛ, a/ ($p > .88$). Figure 2 shows the distribution of the sound duration for the vowels investigated.

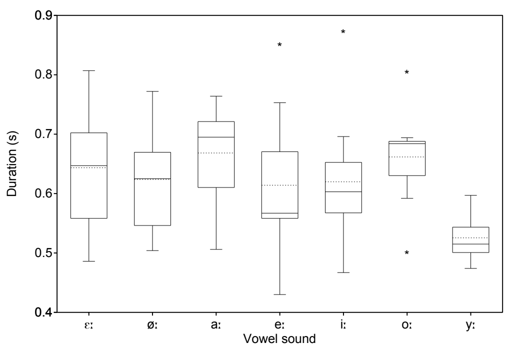

Figure 2: Boxplots representing the distribution of the duration of the vowel sounds.

4. DISCUSSION

The vowels /i, y, o, a/ were consistently identified with ID rates > 80% up to F_0 of 880 Hz. Since F_1 of the speaker at F_0 of 220 Hz corresponds well with statistical F_1 in citation-form words, since /i/ and /y/ are related to the lowest levels of corresponding F_1, and since F_0 of 880 Hz corresponds to the highest F_1 values for all vowels investigated (see Table 1), the results indicate that 'oversinging' statistical F_1 does neither inevitably compromise vowel perception, nor does the perceived vowel category inevitably shift to the category with the next higher F_1. Contrarily, a consistent vowel perception can be maintained independent of statistical F_1. With the exception of the case of /ø/ at F_0 = 587 Hz (possibly due to production inconsistency), the finding that all vowels were identified with a rate > 80% up to F_0 of 740 Hz strongly supports such a conclusion.

Concerning the decrease of the identification rates for /e, ø, ɛ/ above all for $F_0 \geq 740$ Hz, a tendency of a shift in the perceived vowel category to the one with the next higher F_1 is indicated by the results. However, the tendency is inconsistent, i.e., it was not found for all combinations of vowels and F_0-level $\geq$ 740 Hz. Moreover, an alternative interpretation also has to be considered. Discussing possible confusions in terms of shifts towards non-intended vowel categories may have to account for the entire formant patterns of the vowels under investigation and the respective 'density' of neighboring vowel categories according to their placement in the vowel quadrilateral. This would explain why, in the present study, (i) corner vowels were identified more correctly than non-corner vowels, (ii) no strong confusion was found for /o/ (the only back vowel) but some pronounced confusions were found for /e, ø, ɛ/ (three of five front vowels; note also that F_1 is comparable for /o/ and /e, ø/), (iii) identification of /i, y/ (closed front vowels) proved to be more consistent than of /e, ø/ (closed-mid front vowels) and of /ɛ/ (open-mid front vowel); (iv) perceptual shifts were not limited towards the vowel category with a higher F_1 (see Table 2, vowels and F_0 without clear shift tendencies, and the shift /ø/-/e/ at F_0 of 831 Hz).

Statistical analysis does not indicate a clear relation between sound duration and identification performance of the vowels. In line with earlier studies of possible vowel identification at high pitches, the present investigation again shows that vowel perception at very different levels of F_0 cannot be directly related to vowel-specific formant patterns as given for citation-form words. Although a high vocal ability of the speaker and a modified vowel production (e.g., raised larynx, adaption of articulation) may play a crucial role for the present findings, and in addition, dynamic spectral characteristics because of the consonantal context (see e.g., [17, 18]; however, for controversial position, see [5]), and meaning of the /l-V-gen/ tokens may also have a substantial impact on vowel perception, these factors do not allow for a satisfactory explanation concerning the acoustic cues listeners referred to when perceiving the vowels at the very different levels of F_0. Thus, the acoustic cues of vowel perception including all F_0 of vowel identifiability are still a matter of investigation for future research.

5. ACKNOWLEDGEMENTS

This work was supported by the Forschungskredit of the University of Zurich, grant FK-14-062, and the Swiss National Science Foundation (SNSF), grant 100016_143943/1.

6. REFERENCES

[1] Benolken, M.S., Swanson, C.E. 1990. The effect of pitch-related changes on the perception of sung vowels. *J. Acoust. Soc. Am.* 87(4), 1781–1785.

[2] Boersma, P. 1993. Accurate short-term analysis of the fundamenta frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences* 17, 97–110.

[3] Boersma, P., Weenink, D. 2014. Praat: doing phonetics by computer [Computer program]. Version 5.4.04, retrieved 29 December 2014 from http://www.praat.org/

[4] Deme, A. 2014. Intelligibility of sung vowels: the effect of consonantal context and the onset of voicing. *J. Voice* 28, 523.e19–25.

[5] Diehl, R. L., McCusker, S. B., Chapman, L. S. (1981). Perceiving vowels in isolation and in consonantal context. *J. Acoust. Soc. Am.* 69(1), 239-248.

[6] Gottfried, T.L., Chew, S.L. 1986. Intelligibility of vowels sung by a countertenor. *J. Acoust. Soc. Am.* 79(1), 124–130.

[7] Gregg, J.W., Scherer, R.C. 2006. Vowel intelligibility in classical singing. *J. Voice* 20, 198–210.

[8] Helmholtz, H. von, 1870. *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*. Braunschweig: Vieweg.

[9] Hillenbrand, J., Getty, L.A., Clark, M.J., Wheeler, K. 1995. Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.* 97(5), 3099–3111.

[10] Hollien, H., Mendes-Schwartz, A.P., Nielsen, K. 2000. Perceptual confusions of high-pitched sung vowels. *J. Voice* 14, 287–298.

[11] Howie, J., Delattre, P. 1962. An experimental study of the effect of pitch on the intelligibility of vowels. *Natl. Assoc. Teach. Sing*., 385–394.

[12] Maurer, D., Mok, P., Friedrichs, D., Dellwo, V. 2014. Intelligibility of high-pitched vowel sounds in the singing and speaking of a female Cantonese Opera singer. *Fifteenth Annu. Conf. Int. Speech Commun. Assoc.* Singapore, 2132–2133.

[13] Maurer, D., Landis, T. 1996. Intelligibility and spectral differences in high pitched vowels. *Folia Phoniatr. Logop.* 48, 1–10.

[14] Morozov, V.P. 1965. Intelligibility in Singing as a Function of Fundamental Voice Pitch. *Sov. Physics–Acoustics* (Translated from *Akust. Zhurnal*, Vol. 10, No. 3, pp. 330–334, July–September, 1964), 395–402.

[15] Pätzold, M., Simpson, A. 1997. Acoustic analysis of German vowels in the Kiel Corpus of Read Speech. *Arbeitsberichte des Instituts für Phonetik und Digit. Sprachverarbeitung Univ. Kiel* 32, 215–247.

[16] Scotto Di Carlo, N., Germain, A. 1985. A perceptual study of the influence of pitch on the intelligibility of sung vowels. *Phonetica* 42, 188–197.

[17] Smith, L.A., Scott, B.L. 1980. Increasing the intelligibility of sung vowels. *J. Acoust. Soc. Am.* 67(5), 1795–1797.

[18] Strange, W., Verbrugge, R.R., Shankweiler, D.P., Edman, T.R. 1976. Consonant environment specifies vowel identity. *J. Acoust. Soc. Am.* 60(1), 213–224.

[19] Sundberg, J. 1977. Vibrato and vowel identification. *Arch. Acoust*, 257–266.

[20] Sundberg, J. Gauffin, J. 1982. Amplitude of the voice source fundamental and the intelligibility of super pitch vowels. In: Carlson R., Granström, B. (eds), *The representation of speech in the peripheral auditory system.* Amsterdam: Elsevier Biomedlical Press, 223–228.

[21] Sundberg, J. (1987). *The Science of the Singing Voice*. DeKalb, IL: Northern Illinois University Press.

[22] Sundberg, J. 1994. Perceptual aspects of singing. *J. Voice* 8, 106–122.

[23] Sundberg, J. 2012. Perception of Singing. In: Deutsch, D. (eds), *The Psychology of Music*. London: Academic Press, 69–106.